Central Limit Theorem

For random vectors with IID components having a normal distibution, we considered the following theorem:

Theorem If X is a vector of IID random variables each having a $N(\mu, \sigma)$ distribution, then the **sample mean** defined by

$$\overline{x} \;=\; \frac{1}{n}\sum_{i=1}^{n} X_i$$

has a normal distribution with mean μ and standard deviation $\sigma/\sqrt{n}$

Central Limit Theorem

For random vectors with IID components having a normal distibution, we considered the following theorem:

Theorem If X is a vector of IID random variables each having a $N(\mu, \sigma)$ distribution, then the **sample mean** defined by

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

has a normal distribution with mean μ and standard deviation $\sigma / \sqrt{n}$

This begs the question, what if we have a vector of IID random variables with some other distribution?

Central Limit Theorem

Remarkably, for sufficiently large samples, we can remove the normality assumption:

Theorem If X is a vector of IID random variables with expected value μ and standard deviation σ, then the **sample mean** defined by

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

is approximately normally distributed with mean μ and standard deviation $\sigma/\sqrt{n}$

Central Limit Theorem

Remarkably, for sufficiently large samples, we can remove the normality assumption:

Theorem If X is a vector of IID random variables with expected value μ and standard deviation σ, then the **sample mean** defined by

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

is approximately normally distributed with mean μ and standard deviation $\sigma/\sqrt{n}$

This result is a version of an important result called the **central limit theorem**

Central Limit Theorem

The central limit theorem greatly simplifies the practice of statistics, because it allows us to treat means as normally distributed random variables, even if the underlying population is not normal.

Central Limit Theorem

The central limit theorem greatly simplifies the practice of statistics, because it allows us to treat means as normally distributed random variables, even if the underlying population is not normal.

This is very important, because as long as we can take a sufficiently large sample, we don't have to be concerned with the exact form of the underlying population distribution.

Central Limit Theorem

The central limit theorem greatly simplifies the practice of statistics, because it allows us to treat means as normally distributed random variables, even if the underlying population is not normal.

This is very important, because as long as we can take a sufficiently large sample, we don't have to be concerned with the exact form of the underlying population distribution.

In real life, we would prefer not to have to make any assumptions about this underlying distribution, because we can seldom be sure they are met.

Central Limit Theorem

Example A sample of size 100 may be considered an IID random vector with mean $\mu = 10$ and standard deviation $\sigma = 5$.

According to the central limit theorem, what is the approximate distribution of the sample mean?

Central Limit Theorem

Example A sample of size 100 may be considered an IID random vector with mean $\mu = 10$ and standard deviation $\sigma = 5$.

According to the central limit theorem, what is the approximate distribution of the sample mean?

The central limit theorem states that in this case,

$$\overline{x} \approx N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \; = \; N\left(10, \frac{5}{\sqrt{100}}\right) \; = \; N(10, 0.5)$$

Central Limit Theorem

Example A random sample of size 200 is drawn from a population with mean $\mu = 50$ and standard deviation $\sigma = 10$.

Use the central limit theorem to find the approximate probability that $\overline{x} < 51$.

Central Limit Theorem

Example A random sample of size 200 is drawn from a population with mean $\mu = 50$ and standard deviation $\sigma = 10$.

Use the central limit theorem to find the approximate probability that $\overline{x} < 51$.

The central limit theorem states that in this case,

$$\overline{x} \approx N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \;=\; N\left(50, \frac{10}{\sqrt{200}}\right)$$

Then $P(\overline{x} < 51)$ can be computed as:

`pnorm(51,50,10/sqrt(200))` or 0.9213504

Central Limit Theorem

Example A random sample of size 150 is drawn from a population with mean $\mu = -30$ and standard deviation $\sigma = 20$.

Use the central limit theorem to find the approximate probability that $\overline{x} < -25$.

Central Limit Theorem

Example A random sample of size 150 is drawn from a population with mean $\mu = -30$ and standard deviation $\sigma = 20$.

Use the central limit theorem to find the approximate probability that $\overline{x} < -25$.

The central limit theorem states that in this case,

$$\overline{x} \approx N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \;=\; N\left(-30, \frac{20}{\sqrt{150}}\right)$$

Then $P(\overline{x} < -25)$ can be computed as:

`pnorm(-25,-30,20/sqrt(150))` or 0.9989002

Central Limit Theorem

Example A random sample of size 1500 is drawn from a population with mean $\mu = 500$ and standard deviation $\sigma = 100$.

Use the central limit theorem to find the approximate probability that $\overline{x} < 498$.

Central Limit Theorem

Example A random sample of size 1500 is drawn from a population with mean $\mu = 500$ and standard deviation $\sigma = 100$.

Use the central limit theorem to find the approximate probability that $\overline{x} < 498$.

The central limit theorem states that in this case,

$$\overline{x} \approx N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \;=\; N\left(500, \frac{100}{\sqrt{1500}}\right)$$

Then $P(\overline{x} < 498)$ can be computed as:

`pnorm(498,500,100/sqrt(1500))` or 0.219289

Central Limit Theorem

Example A random sample of size 250 is drawn from a population with mean $\mu = 100$ and standard deviation $\sigma = 15$.

Use the central limit theorem to find the approximate probability that $\overline{x}$ is between 100 and 101.

Central Limit Theorem

Example A random sample of size 250 is drawn from a population with mean $\mu = 100$ and standard deviation $\sigma = 15$.

Use the central limit theorem to find the approximate probability that $\overline{x}$ is between 100 and 101.

The central limit theorem states that in this case,

$$\overline{x} \approx N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \; = \; N\left(100, \frac{15}{\sqrt{250}}\right)$$

Then $P(100 < \overline{x} < 101)$ can be computed as:

`pnorm(101,100,15/sqrt(250))`
`-pnorm(100,100,15/sqrt(250))` or 0.3540797